

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/57, 9/64, 5/10, C07K 16/40,</b> <b>C12Q 1/68, A61K 38/43</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 99/58692</b> <b>(43) International Publication Date:</b> 18 November 1999 (18.11.99)
<b>(21) International Application Number:</b> PCT/US99/10386 <b>(22) International Filing Date:</b> 11 May 1999 (11.05.99) <b>(30) Priority Data:</b> 09/078,402 13 May 1998 (13.05.98) US <b>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application</b> US 09/078,402 (CIP) Filed on 13 May 1998 (13.05.98) <b>(71) Applicant (for all designated States except US):</b> INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94034 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). PATTERSON, Chandra [US/US]; 2189 Leland Avenue, Mountain View, CA 94040 (US). BAUGHN,	<b>Mariah [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).</b> <b>(74) Agents:</b> BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US). <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
<b>(54) Title:</b> HUMAN APOPTOSIS ASSOCIATED PROTEINS  <b>(57) Abstract</b>  The invention provides human apoptosis associated proteins (HAPOP) and polynucleotides which identify and encode HAPOP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of HAPOP.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## HUMAN APOPTOSIS ASSOCIATED PROTEINS

### TECHNICAL FIELD

5        This invention relates to nucleic acid and amino acid sequences of human apoptosis associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of diseases associated with increased or decreased apoptosis.

### BACKGROUND OF THE INVENTION

10        Normal development, growth, and homeostasis in multicellular organisms require a careful balance between the production and destruction of cells in tissues throughout the body. Cell division is a carefully coordinated process with numerous checkpoints and control mechanisms. These mechanisms are designed to regulate DNA replication and to prevent inappropriate or excessive cell proliferation. In contrast, apoptosis is the  
15        genetically controlled process by which unneeded or defective cells undergo programmed cell death. Unlike necrotic or injured cells, apoptotic cells are rapidly phagocytosed by neighboring cells or macrophages without leaking their potentially damaging contents into the surrounding tissue or triggering an inflammatory response.

      Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis  
20        include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA into nucleosomal-length units, and expression of novel cell surface components.

      Apoptotic events are part of the normal developmental programs of many  
25        multicellular organisms. Selective elimination of cells is as important for morphogenesis and tissue remodeling as cell proliferation and differentiation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells induce apoptosis in tumor cells or virus-infected cells in vitro. Such activity in vivo may halt the spread of virus or tumor proliferation. In addition, immune cells that  
30        fail to distinguish self molecules from foreign molecules must be eliminated to avoid an autoimmune response.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly, resulting in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known to regulate apoptosis both positively and negatively. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, especially proteases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

The rat ventral prostate (RVP) is a model system for the study of hormone-regulated apoptosis. RVP epithelial cells undergo apoptosis in response to androgen deprivation. Messenger RNA (mRNA) transcripts that are up-regulated in the apoptotic RVP have been identified. (Briehl, M. M. and Miesfeld, R. L. (1991) Mol. Endocrinol. 5:1381-1388.) One such transcript encodes RVP.1, the precise role of which in apoptosis has not been determined. The human homolog of RVP.1, hRVP1, is 89% identical to the rat protein. (Katahira, J. et al. (1997) J. Biol. Chem. 272:26652-26658.) hRVP1 is 220 amino acids in length and contains four transmembrane domains. hRVP1 is highly expressed in the lung, intestine, and liver. Interestingly, hRVP1 functions as a low affinity receptor for the Clostridium perfringens enterotoxin, a causative agent of diarrhea in humans and other animals.

Cytokine-mediated apoptosis plays an important role in hematopoiesis and the immune response. Myeloid cells, which are the stem cell progenitors of macrophages, neutrophils, erythrocytes, and other blood cells, proliferate in response to specific cytokines such as granulocyte/macrophage-colony stimulating factor (GM-CSF) and interleukin-3 (IL-3). When deprived of GM-CSF or IL-3, myeloid cells undergo apoptosis. The murine *requiem* (*req*) gene encodes a putative transcription factor required for this apoptotic response in the myeloid cell line FDCP-1. (Gabig, T. G. et al. (1994) J. Biol. Chem. 269:29515-29519.) The Req protein is 371 amino acids in length and contains a nuclear localization signal, a single *Kruppel*-type zinc finger, an acidic domain, and a cluster of four unique zinc-finger motifs enriched in cysteine and histidine residues involved in metal binding. Expression of *req* is not myeloid- or apoptosis-specific, suggesting that additional factors regulate Req activity in myeloid cell apoptosis.

Tumor necrosis factor (TNF) and related cytokines induce apoptosis in lymphoid cells. (Reviewed in Nagata, S. (1997) Cell 88:355-365.) Binding of TNF to its receptor triggers a signal transduction pathway that results in the activation of a cascade of related proteases, called caspases. One such caspase, ICE (Interleukin-1 $\beta$  converting enzyme), is  
5 a cysteine protease comprised of two large and two small subunits generated by ICE auto-cleavage. (Dinarello, C. A. (1994) FASEB J. 8:1314-1325.) ICE is expressed primarily in monocytes. ICE processes the cytokine precursor, interleukin-1 $\beta$ , into its active form, which plays a central role in acute and chronic inflammation, bone resorption, myelogenous leukemia, and other pathological processes. ICE and related caspases cause  
10 apoptosis when overexpressed in transfected cell lines.

A final step in the apoptotic effector pathway is the fragmentation of nuclear DNA. Recently, a novel factor linking caspase activity to DNA fragmentation has been identified. (Xuesong, L. et al. (1997) Cell 89:175-184.) This factor, DNA fragmentation factor 45 (DFF-45), is proteolytically activated by caspase and is required for DNA  
15 fragmentation. DFF-45 is 331 amino acids in length and exists in the cell as a heterodimer with a second uncharacterized factor. The amino acid sequence of DFF-45 indicates that it is not a nuclease, suggesting that DFF-45 may activate a downstream nuclease. In addition, mRNA encoding a protein related to DFF-45 has been isolated from mouse adipogenic cells. (Danesch, U. et al. (1992) J. Biol. Chem. 267:7185-7193.) Expression  
20 of this mRNA is induced in steroid-treated, differentiating adipocytes. The predicted protein, FSP-27 (fat cell-specific, 27 kilodaltons), is highly basic with a predicted isoelectric point of 10.

Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by  
25 decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast, excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) Science 267:1456-  
30 1462.)

The discovery of new human apoptosis associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful

in the diagnosis, treatment, and prevention of diseases associated with increased or decreased apoptosis.

### SUMMARY OF THE INVENTION

- 5           The invention is based on the discovery of new human apoptosis associated proteins (HAPOP), the polynucleotides encoding HAPOP, and the use of these compositions for the diagnosis, treatment, or prevention of diseases associated with increased or decreased apoptosis. The invention features substantially purified polypeptides, human apoptosis associated proteins, referred to collectively as "HAPOP"
- 10 and individually as "HAPOP-1," "HAPOP-2," "HAPOP-3," and "HAPOP-4." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7.
- 15           The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, or to a fragment of either of these sequences. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3,
- 20 SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ
- 25 ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7.

          Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1,

30 SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7, as well as an isolated and purified polynucleotide having a sequence which is complementary to the

polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7.

- 5           The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8. The invention further provides an isolated and purified polynucleotide variant having at least 90%
- 10 polynucleotide sequence identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8, as well as an isolated and purified polynucleotide having a sequence which is complementary to the
- 15 polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8.

- The invention further provides an expression vector containing at least a fragment
- 20 of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7. In another aspect, the expression vector is contained within a host cell.

- 25           The invention also provides a method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a
- 30 fragment of a polynucleotide encoding the polypeptide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment  
5 of SEQ ID NO:7 in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7, as well as a  
10 purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with increased or decreased apoptosis, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having an amino acid sequence selected  
15 from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7.

The invention also provides a method for treating or preventing a disorder associated with increased or decreased apoptosis, the method comprising administering to  
20 a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7.

The invention also provides a method for detecting a polynucleotide encoding the  
25 polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7 in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide  
30 comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5 and a fragment of SEQ ID NO:7 to at least



one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological  
5 sample are amplified by the polymerase chain reaction prior to the hybridizing step.

### BRIEF DESCRIPTION OF THE FIGURES

Figures 1A and 1B show the amino acid sequence alignments among HAPOP-2 (642272; SEQ ID NO:3), mouse FSP-27 (SWISS-PROT P56198; SEQ ID NO:9), and  
10 human DFF-45 (GI 2065561; SEQ ID NO:10).

Figures 2A and 2B show the amino acid sequence alignment between HAPOP-3 (1453807; SEQ ID NO:5) and mouse Req (GI 606661; SEQ ID NO:11).

Figure 3 shows the amino acid sequence alignment between HAPOP-4 (2059022; SEQ ID NO:7) and human hRVP1 (GI 2570129; SEQ ID NO:12).

15 These alignments were produced using the multisequence alignment program of LASERGENE™ software (DNASTAR Inc, Madison WI).

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is  
20 understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

25 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

30 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those

described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in  
5 connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

#### DEFINITIONS

"HAPOP," as used herein, refers to the amino acid sequences of substantially  
10 purified HAPOP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to HAPOP, increases or prolongs the duration of the effect of HAPOP. Agonists may  
15 include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HAPOP.

An "allelic variant," as this term is used herein, is an alternative form of the gene encoding HAPOP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or  
20 function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

25 "Altered" nucleic acid sequences encoding HAPOP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as HAPOP or a polypeptide with at least one functional characteristic of HAPOP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the  
30 polynucleotide encoding HAPOP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HAPOP. The encoded protein may also be "altered," and may contain

deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HAPOP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HAPOP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

10 The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of HAPOP which are preferably about 5 to about 15 amino acids in length, most preferably 14 amino acids, and which  
15 retain some biological activity or immunological activity of HAPOP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

20 "Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

25 The term "antagonist," as it is used herein, refers to a molecule which, when bound to HAPOP, decreases the amount or the duration of the effect of the biological or immunological activity of HAPOP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HAPOP.

As used herein, the term "antibody" refers to intact molecules as well as to  
30 fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HAPOP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the

immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and  
5 keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the  
10 protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a  
15 nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the  
20 designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HAPOP, or of any oligopeptide thereof, to induce a specific immune response in  
25 appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial,"  
30 such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and

strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding HAPOP or fragments of HAPOP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts, e.g., NaCl, detergents, e.g., sodium dodecyl sulfate (SDS), and other components, e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.

"Consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding HAPOP, by Northern analysis is indicative of the presence of nucleic acids encoding HAPOP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HAPOP.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at

least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5       The term "similarity," as used herein, refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary  
10       sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency  
15       are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or  
20       probe will not hybridize to the second non-complementary target sequence.

      The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign™ program (DNASTAR, Inc., Madison WI). The MegAlign™ program can create alignments  
25       between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of  
30       sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino

acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods  
5 known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) *Nat Genet.* 15:345-355.)

10 The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand  
15 of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another  
20 nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or  
25 nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

30 The term "microarray," as used herein, refers to an arrangement of distinct polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms "element" or "array element" as used herein in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate," as it appears herein, refers to a change in the activity of HAPOP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HAPOP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to



extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding HAPOP, or fragments thereof, or  
5 HAPOP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The  
10 interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

15 As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent (e.g., formamide), temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the  
20 concentration of formamide, or raising the hybridization temperature.

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent,  
25 e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g.,  
30 sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization

will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection,

electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or  
5 RNA for limited periods of time.

A "variant" of HAPOP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have  
10 "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE™ software.

15

#### THE INVENTION

The invention is based on the discovery of new human apoptosis associated proteins (HAPOP), the polynucleotides encoding HAPOP, and the use of these compositions for the diagnosis, treatment, or prevention of diseases associated with  
20 increased or decreased apoptosis.

Nucleic acids encoding the HAPOP-1 of the present invention were first identified in Incyte Clone 157658 from the promonocyte cDNA library (THP1PLB02) using a computer search, e.g., BLAST, for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended  
25 nucleic acid sequences: Incyte Clones 157658 (THP1PLB02), 2451287 (ENDANOT01), and 1241829 (LUNGNOT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. HAPOP-1 is 480 amino acids in length and has one potential N-glycosylation site at N262; eight potential casein kinase II  
30 phosphorylation sites at S51, S130, S193, S227, S238, S371, S411, and S458; and five potential protein kinase C phosphorylation sites at S181, S193, S428, S458, and T470. Both BLOCKS and PRINTS analyses indicate that regions of HAPOP-1 show similarity

to conserved motifs in ICE. These regions extend from P250 to R303, from F308 to G316, and from S344 to S363. A region of unique sequence in HAPOP-1 from about amino acid 203 to about amino acid 212 is encoded by a fragment of SEQ ID NO:2 from about nucleotide 1218 to about nucleotide 1247. Northern analysis shows the expression  
5 of this sequence in various libraries, at least 61% of which are associated with proliferating or cancerous tissue and at least 35% of which are associated with the immune response. In particular, 26% of the libraries expressing HAPOP-1 are derived from reproductive tissue.

Nucleic acids encoding the HAPOP-2 of the present invention were first identified in Incyte Clone 642272 from the breast cDNA library (BRSTNOT03) using a computer  
10 search, e.g., BLAST, for amino acid sequence alignments. A consensus sequence, SEQ ID NO:4, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 642272 (BRSTNOT03), 1954870 (CONNNOT01), 2904084 (DRGCNOT01), and 1965381 (BRSTNOT04).

In one embodiment, the invention encompasses a polypeptide comprising the  
15 amino acid sequence of SEQ ID NO:3. HAPOP-2 is 238 amino acids in length and has four potential casein kinase II phosphorylation sites at S48, S62, T90, and T164 and four potential protein kinase C phosphorylation sites at S21, S36, S53, and S131. As shown in Figures 1A and 1B, HAPOP-2 has chemical and structural similarity with mouse FSP-27 (SWISS-PROT P56198; SEQ ID NO:9) and human DFF-45 (GI 2065561; SEQ ID  
20 NO:10). In particular, HAPOP-2 and FSP-27 share 79% identity, and HAPOP-2 and DFF-45 share 18% identity. In addition, the potential phosphorylation sites at S48, S62, T164, and S131 in HAPOP-2 are conserved in FSP-27, and the potential phosphorylation site at T90 in HAPOP-2 is conserved in both FSP-27 and DFF-45. Like FSP-27, HAPOP-2 is a basic protein with a predicted isoelectric point of 8.8. A region of unique sequence  
25 in HAPOP-2 from about amino acid 31 to about amino acid 40 is encoded by a fragment of SEQ ID NO:4 from about nucleotide 229 to about nucleotide 258. Northern analysis shows the expression of this sequence in various libraries, at least 55% of which are associated with proliferating or cancerous tissue and at least 48% of which are associated with the immune response. In particular, 35% of the libraries expressing HAPOP-2 are  
30 derived from reproductive tissue, and 30% are derived from gastrointestinal tissue.

Nucleic acids encoding the HAPOP-3 of the present invention were first identified in Incyte Clone 1453807 from the penis tumor cDNA library (PENITUT01) using a

computer search, e.g., BLAST, for amino acid sequence alignments. A consensus sequence, SEQ ID NO:6, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1453807 (PENITUT01), 2690812 (LUNGNOT23), 1392458 (THYRNOT03), 2344196 (TESTTUT02), 1995291 and 899939 (BRSTTUT03),  
5 1495966 (PROSNON01), 2083186 (UTRSNOT08), and 3239041 (COLAUCT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:5. HAPOP-3 is 410 amino acids in length and has one potential N-glycosylation site at N237; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S182; eight potential casein kinase II  
10 phosphorylation sites at S28, S182, S209, S239, S243, S247, S314, and S323; eight potential protein kinase C phosphorylation sites at S14, S28, T100, T157, T264, S301, S382, and T396; and two potential tyrosine kinase phosphorylation sites at Y137 and Y188. As shown in Figures 2A and 2B, HAPOP-3 has chemical and structural similarity with mouse Req (GI 606661; SEQ ID NO:11). HAPOP-3 and Req share 20% identity. In  
15 particular, the region of HAPOP-3 from I291 to Q391 is 44% identical to the corresponding region of Req, which comprises the cluster of four unique zinc finger motifs. All 16 cysteine and histidine residues required for metal binding in Req are conserved in HAPOP-3. A region of unique sequence in HAPOP-3 from about amino acid 36 to about amino acid 45 is encoded by a fragment of SEQ ID NO:6 from about  
20 nucleotide 181 to about nucleotide 210. Northern analysis shows the expression of this sequence in various libraries, at least 67% are associated with proliferating or cancerous tissue and at least 33% are associated with the immune response. In particular, 35% of the libraries expressing HAPOP-3 are derived from reproductive tissue, and 27% are derived from gastrointestinal tissue.

25 Nucleic acids encoding the HAPOP-4 of the present invention were first identified in Incyte Clone 2059022 from the ovarian cDNA library (OVARNOT03) using a computer search, e.g., BLAST, for amino acid sequence alignments. A consensus sequence, SEQ ID NO:8, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2059022 (OVARNOT03), 1864717 (PROSNOT19),  
30 814715 (OVRTUT01), 772645 (COLNCRT01), 1317994 (BLADNOT04), 1434058 (BEPINON01), and 1222796 (COLNTUT02).

In one embodiment, the invention encompasses a polypeptide comprising the

amino acid sequence of SEQ ID NO:12. HAPOP-4 is 211 amino acids in length and has two potential N-glycosylation sites at N189 and N205; one potential casein kinase II phosphorylation site at S206; two potential protein kinase C phosphorylation sites at S63 and S206; and a potential signal peptide sequence from M1 to about C24. As shown in

5 Figure 3, HAPOP-4 has chemical and structural similarity with hRVP1 (GI 2570129; SEQ ID NO:12). In particular, HAPOP-4 and hRVP1 share 44% identity. Hydrophobicity plots demonstrate that the four transmembrane domains of hRVP1 are well conserved in HAPOP-4 from about amino acid 1 to about amino acid 29; from about amino acid 78 to about amino acid 103; from about amino acid 120 to about amino acid 140; and from

10 amino acid 162 to about amino acid 186. A region of unique sequence in HAPOP-4 from about amino acid 30 to about amino acid 39 is encoded by a fragment of SEQ ID NO:8 from about nucleotide 520 to about nucleotide 549. Northern analysis shows the expression of this sequence in various libraries, at least 70% are associated with proliferating or cancerous tissue and at least 30% are associated with the immune

15 response. In particular, 38% of the libraries expressing HAPOP-4 are derived from reproductive tissue, and 24% are derived from gastrointestinal tissue.

The invention also encompasses HAPOP variants. A preferred HAPOP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HAPOP amino acid sequence, and

20 which contains at least one functional or structural characteristic of HAPOP.

The invention also encompasses polynucleotides which encode HAPOP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes HAPOP-1. In a further embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID

25 NO:4, which encodes HAPOP-2. In another embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:6, which encodes HAPOP-3. In still another embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:8, which encodes HAPOP-4.

The invention also encompasses a variant of a polynucleotide sequence encoding

30 HAPOP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HAPOP. A

particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. The invention further encompasses a polynucleotide variant of SEQ ID NO:4 having at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:4. The invention further encompasses a polynucleotide variant of SEQ ID NO:6 having at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:6. The invention further encompasses a polynucleotide variant of SEQ ID NO:8 having at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:8. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HAPOP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HAPOP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HAPOP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HAPOP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HAPOP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HAPOP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HAPOP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally

occurring sequence.

The invention also encompasses production of DNA sequences which encode HAPOP and HAPOP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available  
5 expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HAPOP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown  
10 in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, or a fragment of SEQ ID NO:8, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.)

15 Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp., Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading  
20 exonucleases such as those found in the ELONGASE™ Amplification System (GIBCO BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

25 The nucleic acid sequences encoding HAPOP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g.,  
30 Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a



known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this  
5 method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk  
10 genomic DNA (Clontech, Palo Alto, CA). This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO™ 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of  
15 about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be  
20 useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a  
25 charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA  
30 fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HAPOP may be cloned in recombinant DNA molecules that direct

expression of HAPOP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HAPOP.

5       The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HAPOP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the  
10   nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

      In another embodiment, sequences encoding HAPOP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al.  
15   (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, HAPOP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science  
269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide  
20   Synthesizer (Perkin Elmer). Additionally, the amino acid sequence of HAPOP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

      The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol.  
25   182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.)

      In order to express a biologically active HAPOP, the nucleotide sequences encoding HAPOP or derivatives thereof may be inserted into an appropriate expression  
30   vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and

5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HAPOP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HAPOP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HAPOP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HAPOP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HAPOP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HAPOP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences

encoding HAPOP can be achieved using a multifunctional E. coli vector such as Bluescript® (Stratagene) or pSport1™ plasmid (GIBCO BRL). Ligation of sequences encoding HAPOP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing

5 recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of HAPOP are needed, e.g. for the production of antibodies, vectors which direct high level expression of HAPOP may be

10 used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HAPOP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris.

15 In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-54; Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of HAPOP. Transcription of

20 sequences encoding HAPOP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science

25 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized.

30 In cases where an adenovirus is used as an expression vector, sequences encoding HAPOP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the

viral genome may be used to obtain infective virus which expresses HAPOP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may  
5 also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

10 For long term production of recombinant proteins in mammalian systems, stable expression of HAPOP in cell lines is preferred. For example, sequences encoding HAPOP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be  
15 allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

20 Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* or *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for  
25 selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g.,  
30 *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP) (Clontech, Palo Alto, CA),  $\beta$

glucuronidase and its substrate  $\beta$ -D-glucuronoside, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) *Methods Mol. Biol.* 55:121-131.)

5        Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HAPOP is inserted within a marker gene sequence, transformed cells containing sequences encoding HAPOP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a  
10       sequence encoding HAPOP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

         In general, host cells that contain the nucleic acid sequence encoding HAPOP and that express HAPOP may be identified by a variety of procedures known to those of skill  
15       in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

         Immunological methods for detecting and measuring the expression of HAPOP  
20       using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HAPOP is preferred, but a competitive binding assay may be  
25       employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; Coligan, J. E. et al. (1997 and periodic supplements) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; and Maddox, D.E. et al. (1983) *J. Exp. Med.* 158:1211-1216).

30       A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to

polynucleotides encoding HAPOP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HAPOP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HAPOP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HAPOP may be designed to contain signal sequences which direct secretion of HAPOP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HAPOP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HAPOP protein containing a heterologous moiety that can be recognized by a

commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HAPOP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HAPOP encoding sequence and the heterologous protein sequence, so that HAPOP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10. A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HAPOP may be achieved in vitro using the TNT™ rabbit reticulocyte lysate or wheat germ extract systems (Promega, Madison, WI). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably <sup>35</sup>S-methionine.

Fragments of HAPOP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of HAPOP may be synthesized separately and then combined to produce the full length molecule.

#### THERAPEUTICS

Chemical and structural similarity exists between HAPOP-1 and conserved motifs



in ICE. In addition, HAPOP-1 is expressed in proliferating, immune, and reproductive tissue.

Chemical and structural similarity exists among HAPOP-2, mouse FSP-27 (SWISS-PROT P56198), and human DFF-45 (GI 2065561). Chemical and structural  
5 similarity exists between HAPOP-3 and mouse Req (GI 606661) and between HAPOP-4 and human hRVP1 (GI 2570129). In addition, HAPOP-2, HAPOP-3, and HAPOP-4 are expressed in proliferating, immune, reproductive and gastrointestinal tissue.

Therefore, HAPOP appears to play a role in disorders associated with increased or decreased apoptosis, particularly those disorders associated with cell proliferation and the  
10 immune, reproductive and gastrointestinal systems.

Therefore, in one embodiment, HAPOP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with increased or decreased apoptosis. Such disorders can include, but are not limited to, cell proliferative disorders such as atherosclerosis, arteriosclerosis, and cancers such as adenocarcinoma,  
15 leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; immune disorders such as infections, rheumatoid arthritis, dermatitis,  
20 inflammation, systemic lupus erythematosus and other autoimmune diseases; reproductive disorders such as endometriosis, endometrial and ovarian tumors, breast cancer, fibrocystic breast disease, testicular cancer, prostate cancer, and gynecomastia; and gastrointestinal disorders such as esophagitis, esophageal carcinoma, gastritis, gastric carcinoma, inflammatory bowel disease, cholecystitis, infections of the intestinal tract, pancreatitis,  
25 pancreatic carcinoma, cirrhosis, hepatitis, hepatoma, colitis, colonic carcinoma, and Crohn's disease.

In another embodiment, a vector capable of expressing HAPOP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with increased or decreased apoptosis, including, but not limited to, those described above.

30 In a further embodiment, a pharmaceutical composition comprising a substantially purified HAPOP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with increased or

decreased apoptosis, including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HAPOP may be administered to a subject to treat or prevent a disorder associated with increased or decreased apoptosis, including, but not limited to, those listed above.

5 In a further embodiment, an antagonist of HAPOP may be administered to a subject to treat or prevent a disorder associated with increased or decreased apoptosis. Such disorders may include, but are not limited to, those discussed above. In one aspect, an antibody which specifically binds HAPOP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells  
10 or tissue which express HAPOP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HAPOP may be administered to a subject to treat or prevent a disorder associated with increased or decreased apoptosis, including, but not limited to, those described above.

15 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act  
20 synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HAPOP may be produced using methods which are generally known in the art. In particular, purified HAPOP may be used to produce antibodies or to  
25 screen libraries of pharmaceutical agents to identify those which specifically bind HAPOP. Antibodies to HAPOP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially  
30 preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HAPOP or with any fragment or

oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil  
5 emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HAPOP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that  
10 these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HAPOP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

15 Monoclonal antibodies to HAPOP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42;  
20 Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g.,  
25 Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HAPOP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may  
30 be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the

lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HAPOP may also be  
5 generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science  
10 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex  
15 formation between HAPOP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HAPOP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding HAPOP, or  
20 any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HAPOP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HAPOP. Thus, complementary molecules or fragments may be used to modulate HAPOP activity,  
25 or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HAPOP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide  
30 sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding HAPOP. (See, e.g., Sambrook,

supra; and Ausubel, supra.)

Genes encoding HAPOP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HAPOP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HAPOP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HAPOP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene

containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

- 5 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HAPOP. Such DNA
- 10 sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life.

- 15 Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as
- 20 acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

- Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous
- 25 transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

- Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses,
- 30 rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable

carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HAPOP, antibodies to HAPOP, and mimetics, agonists, antagonists, or inhibitors of HAPOP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be  
5 administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular,  
10 intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used  
15 pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral  
20 administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules  
25 (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including  
30 arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with



many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HAPOP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HAPOP or fragments thereof, antibodies of HAPOP, and agonists, antagonists or inhibitors of HAPOP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the  $ED_{50}/LD_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the  
5 subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total  
10 dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations,  
15 etc.

#### DIAGNOSTICS

In another embodiment, antibodies which specifically bind HAPOP may be used for the diagnosis of disorders characterized by expression of HAPOP, or in assays to  
20 monitor patients being treated with HAPOP or agonists, antagonists, or inhibitors of HAPOP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HAPOP include methods which utilize the antibody and a label to detect HAPOP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be  
25 labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HAPOP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels  
30 of HAPOP expression. Normal or standard values for HAPOP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HAPOP under conditions suitable for complex

formation The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HAPOP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing  
5 disease.

In another embodiment of the invention, the polynucleotides encoding HAPOP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in  
10 which expression of HAPOP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HAPOP, and to monitor regulation of HAPOP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HAPOP or closely  
15 related molecules may be used to identify nucleic acid sequences which encode HAPOP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding  
20 HAPOP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the HAPOP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequences of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID  
25 NO:8, or from genomic sequences including promoters, enhancers, and introns of the HAPOP gene.

Means for producing specific hybridization probes for DNAs encoding HAPOP include the cloning of polynucleotide sequences encoding HAPOP or HAPOP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are  
30 commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by

radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HAPOP may be used for the diagnosis of a disorder associated with expression of HAPOP. Examples of such a disorder include, but  
5 are not limited to, a disorder associated with increased or decreased apoptosis, including cell proliferative disorders such as atherosclerosis, arteriosclerosis, and cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle,  
10 ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; immune disorders such as infections, rheumatoid arthritis, dermatitis, inflammation, systemic lupus erythematosus and other autoimmune diseases; reproductive disorders such as endometriosis, endometrial and ovarian tumors, breast cancer, fibrocystic breast disease, testicular cancer, prostate cancer, and gynecomastia; and gastrointestinal  
15 disorders such as esophagitis, esophageal carcinoma, gastritis, gastric carcinoma, inflammatory bowel disease, cholecystitis, infections of the intestinal tract, pancreatitis, pancreatic carcinoma, cirrhosis, hepatitis, hepatoma, colitis, colonic carcinoma, and Crohn's disease. The polynucleotide sequences encoding HAPOP may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR  
20 technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered HAPOP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HAPOP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above.  
25 The nucleotide sequences encoding HAPOP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the  
30 presence of altered levels of nucleotide sequences encoding HAPOP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in

clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HAPOP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal  
5 subjects, either animal or human, with a sequence, or a fragment thereof, encoding HAPOP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with  
10 values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal  
15 subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of  
20 actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HAPOP may involve the use of PCR. These oligomers may be chemically  
25 synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HAPOP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HAPOP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or  
30 quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HAPOP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and

interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a  
5 spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This  
10 information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996)  
15 Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HAPOP  
20 may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries.  
25 (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New  
30 York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HAPOP on a physical chromosomal map and a specific

disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping

5 techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to  
10 investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the  
15 subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HAPOP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening  
20 may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HAPOP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen,  
25 et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with HAPOP, or fragments thereof, and washed. Bound HAPOP is then detected by methods well known in the art. Purified HAPOP can also be coated directly onto plates for use in the aforementioned drug  
30 screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which

neutralizing antibodies capable of binding HAPOP specifically compete with a test compound for binding HAPOP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HAPOP.

In additional embodiments, the nucleotide sequences which encode HAPOP may  
5 be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not  
10 included for the purpose of limiting the invention.

## EXAMPLES

### I. cDNA Library Construction

The BRSTNOT03 cDNA library was constructed from breast tissue removed from  
15 a 54-year-old Caucasian female during a bilateral radical mastectomy. Pathology for the associated tumor tissue indicated residual invasive grade 3 mammary ductal adenocarcinoma. The remaining breast parenchyma exhibited proliferative fibrocystic changes without atypia. One of 10 axillary lymph nodes had metastatic tumor, as a microscopic intranodal focus.

20 The PENITUT01 cDNA library was constructed from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. The tumor involved the glans but did not involve Buck's fascia or corpora cavernosa. Patient history included benign neoplasm of  
25 the large bowel and atherosclerotic coronary artery disease. Family history included malignant neoplasm, chronic lymphocytic leukemia, and chronic liver disease.

The OVARNOT03 cDNA library was constructed from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma.  
30 Patient history included viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, cerebrovascular disease, breast cancer, and uterine cancer.



Frozen tissue from each of the above sources was homogenized and lysed in guanidinium isothiocyanate solution using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NY). The lysate was centrifuged over a CsCl cushion to isolate RNA. The RNA was extracted with phenol, precipitated with sodium acetate  
5 and ethanol, resuspended in RNase-free water, and treated with DNase. The RNA was re-extracted as necessary with acid phenol and reprecipitated.

From each RNA preparation, poly (A+) RNA was isolated using the Qiagen Oligotex kit (QIAGEN, Chatsworth, CA). Poly (A+) RNA was used for cDNA synthesis and construction of each cDNA library according to the recommended protocols in the  
10 SuperScript plasmid system (Catalog #18248-013, Gibco BRL). The cDNAs were size-fractionated on a Sepharose CL4B column (Catalog #275105-01, Pharmacia, Piscataway, NJ). Those cDNAs exceeding 400 bp were ligated into pINCY 1 (Incyte) (PENITUT01) or pSPORT 1 (Gibco BRL) (BRSTNOT03, OVARNOT03). Recombinant plasmids were transformed into DH5 $\alpha$ <sup>TM</sup> competent cells (Catalog #18258-012, Gibco BRL).

15 The THP1PLB02 cDNA library was constructed by reamplification of the THP1PLB01 library, which was custom constructed by Stratagene (La Jolla, CA). THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old male with acute monocytic leukemia. THP-1 cells were activated by culturing in the presence of phorbol ester (PMA) and lipopolysaccharide (LPS). Poly (A+) RNA was isolated from  
20 activated THP-1 cells, and cDNA synthesis was initiated using oligo(dT) and random primers. Double-stranded cDNA was cloned using the Lambda UniZAP vector system (Stratagene).

## II. Isolation and Sequencing of cDNA Clones

### 25 PENITUT01, BRSTNOT03, AND OVARNOT03

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173, QIAGEN Inc) (PENITUT01 and BRSTNOT03) or the Miniprep kit (Catalog #77468, Advanced Genetic Technologies Corporation, Gaithersburg, MD) (OVARNOT03). The recommended protocol was employed except  
30 for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco BRL) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) after the cultures were incubated for 19-24 hours, the cells were lysed with 0.3-0.6 ml of lysis

buffer; and 3) following precipitation, plasmids were stored at 4°C.

#### THP1PLB02

Recombinant pBluescript® phagemid vectors containing cloned cDNAs were recovered by in vivo excision as described by Stratagene. Phagemid DNA was purified  
5 using the QIAwell-8 plasmid purification system (QIAGEN, Chatsworth, CA) or the Magic Minipreps™ DNA purification system (Catalog #A7100, Promega, Madison, WI).

cDNAs from all four libraries were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f). Chain termination products were electrophoresed on urea-  
10 polyacrylamide gels and detected by autoradiography. Alternatively, high-throughput methods for sample preparation utilized the Catalyst 800 or the Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA). cDNAs were sequenced using the Applied Biosystems 373 or 377 DNA sequencing systems in conjunction with fluorescence detection methods; and the  
15 reading frames were determined.

### III. Similarity Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II  
20 databases. These databases, which contain previously identified and annotated sequences, were searched for regions of similarity using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to  
25 determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.)  
30 The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a

database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at  $10^{-25}$  for nucleotides and  $10^{-8}$  for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for  
5 primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for similarity.

Additionally, sequences identified from cDNA libraries may be analyzed to  
10 identify those gene sequences encoding conserved protein motifs using an appropriate analysis program, e.g., BLOCKS. BLOCKS is a weighted matrix analysis algorithm based on short amino acid segments, or blocks, compiled from the PROSITE database. (Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221.) The BLOCKS algorithm is useful for classifying genes with unknown functions. (Henikoff S. And Henikoff G.J.,  
15 *Nucleic Acids Research* (1991) 19:6565-6572.) Blocks, which are 3-60 amino acids in length, correspond to the most highly conserved regions of proteins. The BLOCKS algorithm compares a query sequence with a weighted scoring matrix of blocks in the BLOCKS database. Blocks in the BLOCKS database are calibrated against protein sequences with known functions from the SWISS-PROT database to determine the  
20 stochastic distribution of matches. Similar databases such as PRINTS, a protein fingerprint database, are also searchable using the BLOCKS algorithm. (Attwood, T. K. et al. (1997) *J. Chem. Inf. Comput. Sci.* 37:417-424.) PRINTS is based on non-redundant sequences obtained from sources such as SWISS-PROT, GenBank, PIR, and NRL-3D.

The BLOCKS algorithm searches for matches between a query sequence and the  
25 BLOCKS or PRINTS database and evaluates the statistical significance of any matches found. Matches from a BLOCKS or PRINTS search can be evaluated on two levels, local similarity and global similarity. The degree of local similarity is measured by scores, and the extent of global similarity is measured by score ranking and probability values. A score of 1000 or greater for a BLOCKS match of highest ranking indicates that the match  
30 falls within the 0.5 percentile level of false positives when the matched block is calibrated against SWISS-PROT. Likewise, a probability value of less than  $1.0 \times 10^{-3}$  indicates that the match would occur by chance no more than one time in every 1000 searches. Only

those matches with a cutoff score of 1000 or greater and a cutoff probability value of  $1.0 \times 10^{-3}$  or less are considered in the functional analyses of the protein sequences in the Sequence Listing.

Nucleic and amino acid sequences of the Sequence Listing may also be analyzed using PFAM. PFAM is a Hidden Markov Model (HMM) based protocol useful in protein family searching. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) *Cur. Opin. Str. Biol.* 6:361-365.)

The PFAM database contains protein sequences of 527 protein families gathered from publicly available sources, e.g., SWISS-PROT and PROSITE. PFAM searches for well characterized protein domain families using two high-quality alignment routines, seed alignment and full alignment. (See, e.g., Sonnhammer, E.L.L. et al. (1997) *Proteins* 28:405-420.) The seed alignment utilizes the hmmls program, a program that searches for local matches, and a non-redundant set of the PFAM database. The full alignment utilizes the hmmfs program, a program that searches for multiple fragments in long sequences, e.g., repeats and motifs, and all sequences in the PFAM database. A result or score of 100 "bits" can signify that it is  $2^{100}$ -fold more likely that the sequence is a true match to the model or comparison sequence. Cutoff scores which range from 10 to 50 bits are generally used for individual protein families using the SWISS-PROT sequences as model or comparison sequences.

Two other algorithms, SIGPEPT and TM, both based on the HMM algorithm described above (see, e.g., Eddy, *supra*; and Sonnhammer, *supra*), identify potential signal sequences and transmembrane domains, respectively. SIGPEPT was created using protein sequences having signal sequence annotations derived from SWISS-PROT. It contains about 1413 non-redundant signal sequences ranging in length from 14 to 36 amino acid residues. TM was created similarly using transmembrane domain annotations. It contains about 453 non-redundant transmembrane sequences encompassing 1579 transmembrane domain segments. Suitable HMM models were constructed using the above sequences and were refined with known SWISS-PROT signal peptide sequences or transmembrane domain sequences until a high correlation coefficient, a measurement of the correctness of the analysis, was obtained. Using the protein sequences from the SWISS-PROT database as a test set, a cutoff score of 11 bits, as determined above, correlated with 91-94% true-positives and about 4.1% false-positives, yielding a correlation coefficient of about 0.87-

0.90 for SIGPEPT. A score of 11 bits for TM will typically give the following results: 75% true positives; 1.72% false positives; and a correlation coefficient of 0.76. Each search evaluates the statistical significance of any matches found and reports only those matches that score at least 11 bits.

5

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, 10 e.g., Sambrook, supra, ch. 7; and Ausubel, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to 15 determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

100

The product score takes into account both the degree of similarity between two sequences 20 and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of Northern analysis are reported as a list of libraries in which the 25 transcript encoding HAPOP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

#### 30 V. Extension of HAPOP Encoding Polynucleotides

The nucleic acid sequences of Incyte Clones 157658, 642272, 1453807, and 2059022 were used to design oligonucleotide primers for extending partial nucleotide

sequences to full length. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide  
 5 sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO™ 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer  
 10 dimerizations was avoided.

Selected human cDNA libraries (GIBCO BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-  
 15 PCR™ kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

20	Step 1	94° C for 1 min (initial denaturation)
	Step 2	65° C for 1 min
	Step 3	68° C for 6 min
	Step 4	94° C for 15 sec
	Step 5	65° C for 1 min
	Step 6	68° C for 7 min
25	Step 7	Repeat steps 4 through 6 for an additional 15 cycles
	Step 8	94° C for 15 sec
	Step 9	65° C for 1 min
	Step 10	68° C for 7:15 min
30	Step 11	Repeat steps 8 through 10 for an additional 12 cycles
	Step 12	72° C for 8 min
	Step 13	4° C (and holding)

A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions  
 35 were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK™ (QIAGEN Inc.), and trimmed of

overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C.

- 5 Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37°C, the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing carbenicillin (2x carb). The following day, several colonies were randomly picked from
- 10 each plate and cultured in 150  $\mu$ l of liquid LB/2x carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5  $\mu$ l from each sample was transferred into a PCR array.

- 15 For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

	Step 1	94° C for 60 sec
20	Step 2	94° C for 20 sec
	Step 3	55° C for 30 sec
	Step 4	72° C for 90 sec
	Step 5	Repeat steps 2 through 4 for an additional 29 cycles
	Step 6	72° C for 180 sec
25	Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

- 30 In like manner, the nucleotide sequences of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

## VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically  
5 described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO™ 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially  
10 purified using a Sephadex™ G-25 superfine size exclusion dextran bead column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, MA).

15 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak,  
20 Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

## VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize  
25 array elements on the surface of a substrate. (See, e.g., Baldeschweiler, *supra*.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed  
30 and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.



Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE™. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; and Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

### VIII. Complementary Polynucleotides

Sequences complementary to the HAPOP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HAPOP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO™ 4.06 software and the coding sequence of HAPOP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HAPOP-encoding transcript.

### IX. Expression of HAPOP

Expression and purification of HAPOP is achieved using bacterial or virus-based expression systems. For expression of HAPOP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HAPOP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG).

Expression of HAPOP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HAPOP by either homologous  
5 recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.  
10 K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HAPOP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein  
15 from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Pharmacia, Piscataway, NJ). Following purification, the GST moiety can be proteolytically cleaved from HAPOP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification  
20 using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak, Rochester, NY). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN Inc, Chatsworth, CA). Methods for protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York,  
25 NY, ch 10, 16. Purified HAPOP obtained by these methods can be used directly in the following activity assay.

#### **X. Demonstration of HAPOP Activity**

An assay for HAPOP activity measures the induction of apoptosis when HAPOP  
30 is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT™ (Life

Technologies, Gaithersburg, MD) and pCR™ 3.1 (Invitrogen, Carlsbad, CA, both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 µg of an additional

5 plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser

10 optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90

15 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G.

20 (1994) Flow Cytometry, Oxford, New York, NY.

#### **XI. Transcriptional Analysis**

The influence of HAPOP on gene expression can also be assessed using highly purified populations of cells transfected with sequences encoding HAPOP and either

25 CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success, NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression

30 of mRNA encoding HAPOP and other genes of interest can be analyzed by Northern analysis or microarray techniques.

## **XII. Production of HAPOP Specific Antibodies**

HAPOP substantially purified using polyacrylamide gel electrophoresis (PAGE)(see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using  
5 standard protocols.

Alternatively, the HAPOP amino acid sequence is analyzed using LASERGENE™ software (DNASTAR Inc.) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of  
10 appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel *supra*, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide  
15 ester (MBS) to increase immunogenicity. (See, e.g., Ausubel *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

20

## **XIII. Purification of Naturally Occurring HAPOP Using Specific Antibodies**

Naturally occurring or recombinant HAPOP is substantially purified by immunoaffinity chromatography using antibodies specific for HAPOP. An immunoaffinity column is constructed by covalently coupling anti-HAPOP antibody to an  
25 activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HAPOP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HAPOP (e.g.,  
30 high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HAPOP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HAPOP is collected.

**XIV. Identification of Molecules Which Interact with HAPOP**

HAPOP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate  
5 molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HAPOP, washed, and any wells with labeled HAPOP complex are assayed. Data obtained using different concentrations of HAPOP are used to calculate values for the number, affinity, and association of HAPOP with the candidate molecules.

Various modifications and variations of the described methods and systems of the  
10 invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in  
15 molecular biology or related fields are intended to be within the scope of the following claims.

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7.
2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.
- 10 3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.
4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
- 15 5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.
- 20 6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 3.
7. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8.
- 25 8. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 7.
- 30 9. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 7.

10. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

11. A host cell comprising the expression vector of claim 10.

5

12. A method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7, the method comprising the steps of:

- 10
- a) culturing the host cell of claim 11 under conditions suitable for the expression of the polypeptide; and
  - b) recovering the polypeptide from the host cell culture.

13. A pharmaceutical composition comprising the polypeptide of claim 1 in  
15 conjunction with a suitable pharmaceutical carrier.

14. A purified antibody which specifically binds to the polypeptide of claim 1.

15. A purified agonist of the polypeptide of claim 1.

20

16. A purified antagonist of the polypeptide of claim 1.

17. A method for treating or preventing a disorder associated with increased or decreased apoptosis, the method comprising administering to a subject in need of such  
25 treatment an effective amount of the pharmaceutical composition of claim 13.

18. A method for treating or preventing a disorder associated with increased or decreased apoptosis, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 16.

30

19. A method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1,

SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7 in a biological sample, the method comprising the steps of:

- 5       (a)     hybridizing the polynucleotide of claim 6 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and
- (b)     detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide encoding the polypeptide in the biological sample.

10

20.     The method of claim 19 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.



1	M E Y A M K S L S L L Y P P K S L S R H V S V R T S V V T Q Q	642272
1	M D Y A M K S L S L L Y P P R S L S R H V A V S T A V V T Q Q	P56198
1	M E - - - - - - - - - - V T G D A G V P E S G E I R T	GI 2065561
31	L L S E P S P K A P R A R P C R V S T A D R S V R K G I M A	642272
31	L V S K P S R E T P R A R P C R V S T A D R K V R K G I M A	P56198
18	L - - - - - - - - - - K P C L L R R N Y S R E Q H G V A	GI 2065561
61	Y S L E D L L L K V R D T L M L A D K - - P F F L V L E E D	642272
61	H S L E D L L L N K V Q D I L K L K D K - - P F S L V L E E D	P56198
37	S C L E D L R S K A C D I L A I D K S L T P V T L V L A E D	GI 2065561
89	G T V E T E E Y F Q A L A G D T V F M V L Q K G Q K W - -	642272
89	G T I V E T E E Y F Q A L A K D T M F M V L L K G Q K W - -	P56198
67	G T I V D D D Y F L C L P S N T K F V A L A S N E K W A Y	GI 2065561
117	- - - - - - - - - - Q P P S E Q G T R H P - L S L S H K P	642272
117	- - - - - - - - - - K P P S E Q R K K R A Q L A L S Q K P	P56198
97	N N S D G G T A W I S Q E S F D V D E T D S G A G L K W K N	GI 2065561
135	A K - - - - - K I D V A R V T F D L Y K L N P Q D F I G - - C L	642272
136	T K - - - - - K I D V A R V T F D L Y K L N P Q D F I G - - C L	P56198
127	V A R Q L K E D L S S I I L - L S E E D L Q M L V D A P C S	GI 2065561

FIGURE 1A

160	N	V	K	A	T	F	Y	D	T	Y	S	-	-	-	-	L	S	Y	D	L	H	C	C	G	A	K	R	I	-	-	642272
161	N	V	K	A	T	L	Y	D	T	Y	S	-	-	-	-	L	S	Y	D	L	H	C	Y	K	A	K	R	I	-	-	P56198
156	D	L	A	Q	E	L	R	Q	S	C	A	T	V	Q	R	L	Q	H	T	L	Q	Q	V	L	D	Q	R	E	V	GI 2065561	
184	-	-	M	K	E	A	F	R	W	A	L	F	S	M	Q	A	T	G	H	V	L	L	-	-	-	-	G	T	S	C	642272
185	-	-	V	K	E	I	V	R	W	T	L	F	S	M	Q	A	T	G	H	M	L	L	-	-	-	-	G	T	S	S	P56198
186	R	Q	S	K	Q	L	L	Q	L	Y	L	Q	A	L	E	K	E	G	S	L	L	S	K	Q	E	E	S	K	A	A	GI 2065561
208	Y	L	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	L	D	A	T	642272	
209	Y	M	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	F	L	D	A	T	P56198
216	F	G	E	E	V	D	A	V	D	T	G	I	S	R	E	T	S	S	D	V	A	L	A	S	H	I	L	T	A	L	GI 2065561
217	E	E	G	Q	P	P	K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	642272	
218	E	E	Q	P	A	K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P56198		
246	R	E	K	Q	A	P	E	L	S	L	S	S	Q	D	L	E	L	V	T	K	E	D	P	K	A	L	A	V	A	L	GI 2065561
224	-	-	-	-	G	K	A	S	S	L	I	P	T	C	L	K	I	L	Q	-	-	-	-	-	-	-	-	-	-	642272	
225	-	-	-	-	A	K	P	S	S	L	L	P	A	C	L	K	M	L	Q	-	-	-	-	-	-	-	-	-	P56198		
276	N	W	D	I	K	K	T	E	T	V	Q	E	A	C	E	R	E	L	A	L	R	L	Q	Q	T	Q	S	L	H	S	GI 2065561
238	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	642272		
239	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P56198		
306	L	R	S	I	S	A	S	K	A	S	P	P	G	D	L	Q	N	P	K	R	A	R	Q	D	P	T	-	-	-	GI 2065561	

FIGURE 1B

1	M	L	Q	E	Q	V	S	E	Y	L	G	V	T	S	F	K	R	K	Y	P	D	L	E	R	R	D	L	S	H	1453807
1	M	-	-	E	Q	C	H	N	Y	N	A	-	-	-	-	R	L	C	A	E	R	S	V	R	-	L	P	-	GI 606661	
30	K	E	K	L	Y	L	R	E	L	N	V	I	T	E	T	Q	C	T	L	G	L	T	A	-	L	R	S	D	E	1453807
21	-	-	-	-	F	L	D	S	Q	T	G	V	A	Q	S	N	C	Y	I	W	M	E	K	R	H	R	G	P	G	GI 606661
58	V	I	D	L	M	I	K	E	Y	P	A	K	H	A	E	Y	S	V	I	L	Q	E	K	E	R	Q	R	I	T	1453807
46	L	A	S	G	Q	L	Y	S	Y	P	A	R	R	-	-	-	-	W	R	K	K	R	R	A	H	P	P	GI 606661		
87	D	H	Y	K	-	E	Y	S	Q	M	Q	Q	N	T	Q	K	V	E	A	S	K	V	P	E	Y	I	K	K	1453807	
69	E	D	P	R	L	S	F	P	S	I	K	P	D	T	D	Q	T	L	K	K	E	G	L	-	-	-	-	-	GI 606661	
115	A	A	K	K	A	A	E	F	N	S	N	L	N	R	E	R	M	E	R	R	A	Y	F	D	L	Q	T	H	1453807	
92	I	S	Q	D	G	S	S	L	E	A	L	L	R	T	D	P	L	E	K	R	G	A	-	-	-	P	D	P	R	GI 606661
144	V	I	Q	V	P	Q	G	K	Y	K	V	L	P	T	E	R	T	K	V	S	S	Y	P	V	A	L	I	P	G	1453807
118	V	D	D	S	L	G	E	F	P	V	-	S	N	S	R	A	R	K	R	I	-	-	-	I	E	P	D	GI 606661		
173	Q	F	Q	E	Y	Y	K	R	Y	S	P	D	E	L	R	Y	L	P	L	N	T	A	L	Y	E	P	P	L	D	1453807
142	D	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GI 606661	
202	P	E	L	P	A	L	D	S	D	G	D	S	D	D	G	E	D	G	R	G	D	E	K	R	K	N	K	G	T	1453807
144	-	-	-	L	D	D	L	D	E	D	Y	E	E	D	R	P	K	R	R	G	K	G	K	S	K	S	K	G	V	GI 606661

FIGURE 2A

231 S D S S S - - - G N V S E G E S P P - - - D S Q E D S F Q 1453807  
 171 S S A R K K L D A S I L E D R D K P Y A C D I C G K R Y K GI 606661  
  
 254 G R - - - - - - - - - Q K S K D K A A T - P R K 1453807  
 200 N R P G L S Y H Y A H S H L A E E F E D K E D S R P P T GI 606661  
  
 268 D G P K R S V L S K S V P G Y K P K V I P N A I C G I C L 1453807  
 229 P V S Q R S E E Q K S K K G P D G L A L P N N Y C D F C L GI 606661  
  
 297 - K G K E S N K K G K A E S L I H C S Q C E N S G H P S C 1453807  
 258 G D S K I N K K T G Q P E L V S C S D C G R S G H P S C GI 606661  
  
 325 L D M T M E L V S M I K T Y P W Q C M E C K T C I I C G Q 1453807  
 287 L Q F T P V M M A A V K T Y R W Q C I E C K C C N L C G T GI 606661  
  
 354 P H H E E M M F C D M C D R G Y H T F C V - - G L G A I 1453807  
 316 S E N D D Q L L F C D D C D R G Y H M Y C L T P S M S E P GI 606661  
  
 381 P S G R W I C D C C Q R A P P T P R K V G R R G K N S K E 1453807  
 345 P E G S W S C H L C L D L L - - - K E K A S I Y Q N Q N S GI 606661  
  
 410 G 1453807  
 371 S GI 606661

FIGURE 2B

1	M	A	N	S	G	L	Q	L	L	G	F	S	M	A	L	L	G	W	V	G	L	V	A	C	T	A	I	P	Q	W	2059022		
2	M	S	M	-	G	L	E	I	T	G	T	A	L	A	V	L	G	W	L	G	T	I	V	C	C	A	L	P	M	W	GI 2570129		
31	Q	M	S	S	Y	A	G	D	N	I	I	T	A	Q	A	M	Y	K	G	L	W	M	D	C	V	T	Q	S	T	G	2059022		
30	R	V	S	A	F	I	G	S	N	I	I	T	S	Q	N	I	W	E	G	L	W	M	N	C	V	V	Q	S	T	G	GI 2570129		
61	M	M	S	C	K	M	Y	D	S	V	L	A	L	S	A	A	L	Q	A	T	R	A	L	M	V	V	S	L	V	L	2059022		
60	Q	M	Q	C	K	V	Y	D	S	L	L	A	L	P	Q	D	L	Q	A	A	R	A	L	I	V	V	A	I	L	L	GI 2570129		
91	G	F	L	A	M	F	V	A	T	M	G	M	K	C	T	R	C	G	G	D	D	K	V	K	K	A	R	I	A	M	2059022		
90	A	A	F	G	L	V	A	L	V	G	A	Q	C	T	N	C	V	Q	D	D	T	A	K	-	A	K	I	T	I	GI 2570129			
121	G	G	G	I	I	F	I	V	A	G	L	A	A	L	V	A	C	S	W	Y	G	H	Q	I	V	T	D	F	Y	N	2059022		
119	V	A	G	V	L	F	L	A	A	L	L	T	L	V	P	V	S	W	S	A	N	T	I	I	R	D	F	Y	N	GI 2570129			
151	P	L	I	P	T	N	I	K	Y	E	F	G	P	A	I	F	I	G	W	A	G	S	A	L	V	I	L	G	G	A	2059022		
149	P	V	V	P	E	A	Q	K	R	E	M	G	A	G	L	Y	V	G	W	A	A	A	L	Q	L	L	G	G	A	GI 2570129			
181	L	L	S	C	S	C	P	G	N	E	-	-	-	-	-	-	-	S	K	A	G	Y	R	A	P	R	S	Y	P	K	S	N	2059022
179	L	L	C	C	S	C	P	P	R	E	K	K	Y	T	A	T	K	V	Y	S	A	P	R	S	T	G	P	G	A	GI 2570129			
206	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	K	E	Y	V	2059022										
209	S	L	G	T	G	Y	D	R	K	D	Y	V	GI 2570129																				

FIGURE 3

## SEQUENCE LISTING

&lt;110&gt; INCYTE PHARMACEUTICALS, INC.

HILLMAN, Jennifer L.

CORLEY, Neil C.

GUEGLER, Karl J.

PATTERSON, Chandra

BAUGHN, Mariah

&lt;120&gt; HUMAN APOPTOSIS ASSOCIATED PROTEINS

&lt;130&gt; PF-0519 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

&lt;150&gt; 09/078,402

&lt;151&gt; 1998-05-13

&lt;160&gt; 12

&lt;170&gt; Perl Program

&lt;210&gt; 1

&lt;211&gt; 480

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 157658

&lt;400&gt; 1

Met	Ser	Ala	Glu	Val	Ile	His	Gln	Val	Glu	Glu	Ala	Leu	Asp	Thr
1			5						10					15
Asp	Glu	Lys	Glu	Met	Leu	Leu	Phe	Leu	Cys	Arg	Asp	Val	Ala	Ile
			20						25					30
Asp	Val	Val	Pro	Pro	Asn	Val	Arg	Asp	Leu	Leu	Asp	Ile	Leu	Arg
			35						40					45
Glu	Arg	Gly	Lys	Leu	Ser	Val	Gly	Asp	Leu	Ala	Glu	Leu	Leu	Tyr
			50						55					60
Arg	Val	Arg	Arg	Phe	Asp	Leu	Leu	Lys	Arg	Ile	Leu	Lys	Met	Asp
			65						70					75
Arg	Lys	Ala	Val	Glu	Thr	His	Leu	Leu	Arg	Asn	Pro	His	Leu	Val
			80						85					90
Ser	Asp	Tyr	Arg	Val	Leu	Met	Ala	Glu	Ile	Gly	Glu	Asp	Leu	Asp
			95						100					105
Lys	Ser	Asp	Val	Ser	Ser	Leu	Ile	Phe	Leu	Met	Lys	Asp	Tyr	Met
			110						115					120
Gly	Arg	Gly	Lys	Ile	Ser	Lys	Glu	Lys	Ser	Phe	Leu	Asp	Leu	Val
			125						130					135
Val	Glu	Leu	Glu	Lys	Leu	Asn	Leu	Val	Ala	Pro	Asp	Gln	Leu	Asp
			140						145					150
Leu	Leu	Glu	Lys	Cys	Leu	Lys	Asn	Ile	His	Arg	Ile	Asp	Leu	Lys
			155						160					165

Thr Lys Ile Gln Lys Tyr Lys Gln Ser Val Gln Gly Ala Gly Thr	170	175	180
Ser Tyr Arg Asn Val Leu Gln Ala Ala Ile Gln Lys Ser Leu Lys	185	190	195
Asp Pro Ser Asn Asn Phe Arg Leu His Asn Gly Arg Ser Lys Glu	200	205	210
Gln Arg Leu Lys Glu Gln Leu Gly Ala Gln Gln Glu Pro Val Lys	215	220	225
Lys Ser Ile Gln Glu Ser Glu Ala Phe Leu Pro Gln Ser Ile Pro	230	235	240
Glu Glu Arg Tyr Lys Met Lys Ser Lys Pro Leu Gly Ile Cys Leu	245	250	255
Ile Ile Asp Cys Ile Gly Asn Glu Thr Glu Leu Leu Arg Asp Thr	260	265	270
Phe Thr Ser Leu Gly Tyr Glu Val Gln Lys Phe Leu His Leu Ser	275	280	285
Met His Gly Ile Ser Gln Ile Leu Gly Gln Phe Ala Cys Met Pro	290	295	300
Glu His Arg Asp Tyr Asp Ser Phe Val Cys Val Leu Val Ser Arg	305	310	315
Gly Gly Ser Gln Ser Val Tyr Gly Val Asp Gln Thr His Ser Gly	320	325	330
Leu Pro Leu His His Ile Arg Arg Met Phe Met Gly Asp Ser Cys	335	340	345
Pro Tyr Leu Ala Gly Lys Pro Lys Met Phe Phe Ile Gln Asn Tyr	350	355	360
Val Val Ser Glu Gly Gln Leu Glu Asp Ser Ser Leu Leu Glu Val	365	370	375
Asp Gly Pro Ala Met Lys Asn Val Glu Phe Lys Ala Gln Lys Arg	380	385	390
Gly Leu Cys Thr Val His Arg Glu Ala Asp Phe Phe Trp Ser Leu	395	400	405
Cys Thr Ala Asp Met Ser Leu Leu Glu Gln Ser His Ser Ser Pro	410	415	420
Ser Leu Tyr Leu Gln Cys Leu Ser Gln Lys Leu Arg Gln Glu Arg	425	430	435
Lys Arg Pro Leu Leu Asp Leu His Ile Glu Leu Asn Gly Tyr Met	440	445	450
Tyr Asp Trp Asn Ser Arg Val Ser Ala Lys Glu Lys Tyr Tyr Val	455	460	465
Trp Leu Gln His Thr Leu Arg Lys Lys Leu Ile Leu Ser Tyr Thr	470	475	480

&lt;210&gt; 2

&lt;211&gt; 2352

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 157658

&lt;400&gt; 2

gaaattgctgc cactgcactc cagcctgggc cacagagcga gactctgtct caaaaaagaa 60  
 ggaaagaaag aaagaaaaaa aaaaacactc gcagtgttta ctctaacgc gtggaacttg 120

```

tgtcgacatc ccccccggt tactgcatac tcagtcacac aagccatagc aggaaacagc 180
gagcttgcag cctcaccgac gagtctcaac taaaaggagc tcccggagct aggggtgggg 240
actcggcctc acacagttag tgccggctat tggacttttg tccagtgaca gctgagacaa 300
caaggaccac gggaggaggt gtaggagaga agcgccgcga acagcgatcg cccagcacca 360
agtccgcttc caggctttcg gtttctttgc ctccatcttg ggtgcgcctt cccggcgctt 420
aggggagcga aggctgaggt ggcagcggca ggagagtccg gccgcgacag gacgaactcc 480
ccactggaa aggattctga aagaaatgaa gtcagccctc agaaatgaag ttgactgcct 540
gctggctttc tgttgactgg cccggagctg tactgcaaga cccttgtagg ctccctagt 600
ctaagagtag gatgtctgct gaagtcaccc atcagggtga agaagcactt gatacagatg 660
agaaggagat gctgctcttt ttgtgccggg atgttgctat agatgtggtt ccacctaata 720
tcagggaact tctggatatt ttacgggaaa gaggtaagct gtctgtcggg gacttggctg 780
aactgctcta cagagttagg cgatttgacc tgcctaaacg tatcttgaag atggacagaa 840
aagctgtgga gaccacactg ctcaggaacc ctcaccttgt ttcggactat agagtgtgta 900
tggcagagat tgggtgaggat ttggataaat ctgatgtgtc ctcattaatt ttcctcatga 960
aggattacat gggccgaggc aagataagca aggagaagag tttcttgagc cttgtggttg 1020
agttggagaa actaaatctg gttgccccag atcaactgga tttattagaa aaatgcctaa 1080
agaacatcca cagaatagac ctgaagacaa aaatccagaa gtacaagcag tctgttcaag 1140
gagcagggac aagttacagg aatgttctcc aagcagcaat ccaaaagagt ctcaaggatc 1200
cttcaaataa cttcaggctc cataatggga gaagtaaaga acaaagactt aaggaacagc 1260
ttggcgctca acaagaacca gtgaagaaat ccattcagga atcagaagct tttttgcctc 1320
agagcatacc tgaagagaga tacaagatga agagcaagcc cctaggaatc tgcctgataa 1380
tcgattgcat tggcaatgag acagagcttc ttcgagacac cttcacttcc ctgggctatg 1440
aagtcagaa attcttgcac ctcagtatgc atggtatata ccagattctt ggccaatttg 1500
cctgtatgcc cgagcaccga gactacgaca gctttgtgtg tgtcctgggtg agccgaggag 1560
gctcccagag tgtgtatggt gtggatcaga ctcactccgg gctccccctg catcacatca 1620
ggaggatggt catgggagat tcatgccctt atctagcagg gaagccaaag atgtttttta 1680
ttcagaacta tgtggtgtca gagggccagc tggaggacag cagcctcttg gaggtggatg 1740
ggccagcgat gaagaatgtg gaattcaagg ctcagaagcg agggctgtgc acagttcacc 1800
gagaagctga cttcttctgg agcctgtgta ctgcggacat gtccctgctg gagcagcttc 1860
acagctcacc atccctgtac ctgcagtgcc tctcccagaa actgagacaa gaaagaaaac 1920
gcccactcct ggatcttcac attgaactca atggctacat gtatgattgg aacagcagag 1980
tttctgcaa ggagaaatat tatgtctggc tgcagcacac tctgagaaag aaacttatcc 2040
tctcctacac ataagaaacc aaaaggctgg gcgtagtggc tcacacctgt gatcccagca 2100
ctttgggagg ccaaggaggg cagatcactt caggtcagga gttcgagacc agcctggcca 2160
acatggtaaa cgctgtccct agtaaaaata caaaaattag ctgggtgtgg gtgtgggtac 2220
ctgtattccc agttacttgg gaggtgagg tgggaggatc ttttgaaccc aggagttcag 2280
ggcatagca tgctgtgatt gtgcctacga atagccactg cataccaacc tgggcaatat 2340
agcaagatcc ca 2352

```

&lt;210&gt; 3

&lt;211&gt; 238

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 642272

&lt;400&gt; 3

```

Met Glu Tyr Ala Met Lys Ser Leu Ser Leu Leu Tyr Pro Lys Ser
 1             5             10             15
Leu Ser Arg His Val Ser Val Arg Thr Ser Val Val Thr Gln Gln
                20             25             30

```



Leu	Leu	Ser	Glu	Pro	Ser	Pro	Lys	Ala	Pro	Arg	Ala	Arg	Pro	Cys	
				35					40					45	
Arg	Val	Ser	Thr	Ala	Asp	Arg	Ser	Val	Arg	Lys	Gly	Ile	Met	Ala	
				50					55					60	
Tyr	Ser	Leu	Glu	Asp	Leu	Leu	Leu	Lys	Val	Arg	Asp	Thr	Leu	Met	
				65					70					75	
Leu	Ala	Asp	Lys	Pro	Phe	Phe	Leu	Val	Leu	Glu	Glu	Asp	Gly	Thr	
				80					85					90	
Thr	Val	Glu	Thr	Glu	Glu	Tyr	Phe	Gln	Ala	Leu	Ala	Gly	Asp	Thr	
				95					100					105	
Val	Phe	Met	Val	Leu	Gln	Lys	Gly	Gln	Lys	Trp	Gln	Pro	Pro	Ser	
				110					115					120	
Glu	Gln	Gly	Thr	Arg	His	Pro	Leu	Ser	Leu	Ser	His	Lys	Pro	Ala	
				125					130					135	
Lys	Lys	Ile	Asp	Val	Ala	Arg	Val	Thr	Phe	Asp	Leu	Tyr	Lys	Leu	
				140					145					150	
Asn	Pro	Gln	Asp	Phe	Ile	Gly	Cys	Leu	Asn	Val	Lys	Ala	Thr	Phe	
				155					160					165	
Tyr	Asp	Thr	Tyr	Ser	Leu	Ser	Tyr	Asp	Leu	His	Cys	Cys	Gly	Ala	
				170					175					180	
Lys	Arg	Ile	Met	Lys	Glu	Ala	Phe	Arg	Trp	Ala	Leu	Phe	Ser	Met	
				185					190					195	
Gln	Ala	Thr	Gly	His	Val	Leu	Leu	Gly	Thr	Ser	Cys	Tyr	Leu	Gln	
				200					205					210	
Gln	Leu	Leu	Asp	Ala	Thr	Glu	Glu	Gly	Gln	Pro	Pro	Lys	Gly	Lys	
				215					220					225	
Ala	Ser	Ser	Leu	Ile	Pro	Thr	Cys	Leu	Lys	Ile	Leu	Gln			
				230					235						

&lt;210&gt; 4

&lt;211&gt; 1284

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 642272

&lt;400&gt; 4

```

aatgttcttt tggccactgt gaagcctcag gaaggggctc ggattgctca aggacccatg 60
ggagagagga ggctttgact gggctgcctg cctgtgaggt ctctggacta gaggtccaac 120
gcagtccagc tgacaaggat ggaatacgcc atgaagtcct ttagccttct ctaccccaag 180
tccctctcca ggcatgtgtc agtgcgtacc tctgtggtga cccagcagct gctgtcggag 240
cccagcccca agggccccag ggcccggccc tgccgcgtaa gcacggcgga tcgaagcgtg 300
aggaagggca tcatggctta cagtcttgag gacctcctcc tcaaggctcc ggacactctg 360
atgctggcag acaagccctt cttcctgggt ctggaggaag atggcacaac tgtagagaca 420
gaagagtact tccaagccct ggcaggggat acagtgttca tggctcctca gaaggggagc 480
aatggcagc ccccatcaga acaggggaca aggcacccac tgtcctctc ccataagcct 540
gccaagaaga ttgatgtggc ccgtgtaacg tttgatctgt acaagctgaa cccacaggac 600
ttcattggct gctgaacgt gaaggcgact ttttatgata catactccct ttcctatgat 660
ctgcactgct gtggggccaa ggcgcatcat aaggaagctt tccgctgggc cctcttcagc 720
atgcaggcca caggccacgt actgcttggc acctcctgtt acctgcagca gctcctcgat 780
gctacggagg aagggcagcc ccccaagggc aaggcctcat cccttatccc gacctgtctg 840
aagatactgc agtgaaagcc caagtccttg gaagctttcc ccagtgaagg actgactggg 900
ggcctcacgc ttaactggta gtgcccacaa gcttggcagc tgtagagccg cgaacctccc 960

```

```

cacacctccc tcaccgcgca ggaccctgag tgaggaggag gagctggaaa cctgggggtgg 1020
gttggccaaa ggagaacctc aagctcctgg cctgatccag ctccttcctg cccaaggcag 1080
cttagcccat ccagactggg cctgaagtct gtccctccat tggcatgaag tctgcccctt 1140
agcaatccgg cctcgcaggc tgtactttca tgggtgctctc taccttctgg ccccatcccc 1200
ggaacattcc tgagtgaatt cgcaagcgca ctagcatgtg atattaggga gtttgaata 1260
aattattgag gctgaaaaaa aaaa 1284

```

&lt;210&gt; 5

&lt;211&gt; 410

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 1453807

&lt;400&gt; 5

```

Met Leu Gln Glu Gln Val Ser Glu Tyr Leu Gly Val Thr Ser Phe
1      5      10      15
Lys Arg Lys Tyr Pro Asp Leu Glu Arg Arg Asp Leu Ser His Lys
20     25     30
Glu Lys Leu Tyr Leu Arg Glu Leu Asn Val Ile Thr Glu Thr Gln
35     40     45
Cys Thr Leu Gly Leu Thr Ala Leu Arg Ser Asp Glu Val Ile Asp
50     55     60
Leu Met Ile Lys Glu Tyr Pro Ala Lys His Ala Glu Tyr Ser Val
65     70     75
Ile Leu Gln Glu Lys Glu Arg Gln Arg Ile Thr Asp His Tyr Lys
80     85     90
Glu Tyr Ser Gln Met Gln Gln Gln Asn Thr Gln Lys Val Glu Ala
95     100    105
Ser Lys Val Pro Glu Tyr Ile Lys Lys Ala Ala Lys Lys Ala Ala
110    115    120
Glu Phe Asn Ser Asn Leu Asn Arg Glu Arg Met Glu Glu Arg Arg
125    130    135
Ala Tyr Phe Asp Leu Gln Thr His Val Ile Gln Val Pro Gln Gly
140    145    150
Lys Tyr Lys Val Leu Pro Thr Glu Arg Thr Lys Val Ser Ser Tyr
155    160    165
Pro Val Ala Leu Ile Pro Gly Gln Phe Gln Glu Tyr Tyr Lys Arg
170    175    180
Tyr Ser Pro Asp Glu Leu Arg Tyr Leu Pro Leu Asn Thr Ala Leu
185    190    195
Tyr Glu Pro Pro Leu Asp Pro Glu Leu Pro Ala Leu Asp Ser Asp
200    205    210
Gly Asp Ser Asp Asp Gly Glu Asp Gly Arg Gly Asp Glu Lys Arg
215    220    225
Lys Asn Lys Gly Thr Ser Asp Ser Ser Ser Gly Asn Val Ser Glu
230    235    240
Gly Glu Ser Pro Pro Asp Ser Gln Glu Asp Ser Phe Gln Gly Arg
245    250    255
Gln Lys Ser Lys Asp Lys Ala Ala Thr Pro Arg Lys Asp Gly Pro
260    265    270
Lys Arg Ser Val Leu Ser Lys Ser Val Pro Gly Tyr Lys Pro Lys
275    280    285

```

Val	Ile	Pro	Asn	Ala	Ile	Cys	Gly	Ile	Cys	Leu	Lys	Gly	Lys	Glu
			290						295					300
Ser	Asn	Lys	Lys	Gly	Lys	Ala	Glu	Ser	Leu	Ile	His	Cys	Ser	Gln
			305						310					315
Cys	Glu	Asn	Ser	Gly	His	Pro	Ser	Cys	Leu	Asp	Met	Thr	Met	Glu
			320						325					330
Leu	Val	Ser	Met	Ile	Lys	Thr	Tyr	Pro	Trp	Gln	Cys	Met	Glu	Cys
			335						340					345
Lys	Thr	Cys	Ile	Ile	Cys	Gly	Gln	Pro	His	His	Glu	Glu	Glu	Met
			350						355					360
Met	Phe	Cys	Asp	Met	Cys	Asp	Arg	Gly	Tyr	His	Thr	Phe	Cys	Val
			365						370					375
Gly	Leu	Gly	Ala	Ile	Pro	Ser	Gly	Arg	Trp	Ile	Cys	Asp	Cys	Cys
			380						385					390
Gln	Arg	Ala	Pro	Pro	Thr	Pro	Arg	Lys	Val	Gly	Arg	Arg	Gly	Lys
			395						400					405
Asn	Ser	Lys	Glu	Gly										
			410											

&lt;210&gt; 6

&lt;211&gt; 2150

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 1453807

&lt;400&gt; 6

```

cttggtttta gttactatcc agcagaaaac ttgatagagt acaaattggcc acctgatgaa 60
acaggagaat actatatgct tcaagaacaa gtcagtgaat atttgggtgt gacctccttt 120
aaaaggaaat atccagattt agagcgacga gatttgtctc acaaggagaa actctacctg 180
agagagctaa atgtcattac tgaaactcag tgcactctag gcttaacagc attgcgagct 240
gatgaagtga ttgatttaat gataaaagaa tatccagcca aacatgctga gtattctgtt 300
attctacaag aaaaagaacg tcaacgaatt acagaccatt ataaagagta ttcccaaagt 360
caacaacaga atactcagaa agttgaagcc agtaaagtgc ctgagtatat taagaaagct 420
gccaaaaaag cagcagaatt taatagcaac ttaaacggg aacgcagtgga agaaagaaga 480
gcttattttg acttgacagc acatgtttatc caggtaacctc aagggaagta caaagttttg 540
ccaacagagc gaacaaaggc cagttcttac ccagtggctc tcatccccgg acagttccag 600
gaatattata agaggtactc accagatgag ctgcggtatc tgccattaaa cacagccctg 660
tatgagcccc ctctggatcc tgagctccct gctctagaca gtgatgggta ttcagatgat 720
ggcgaagatg gtcgaggtga tgagaaacgg aaaaataaag gcacttcgga cagctcctct 780
ggcaatgtat ctgaagggga aagccctcct gacagccagg aggactcttt ccagggaaga 840
cagaaatcaa aagacaaagc tgccactcca agaaaagatg gtcccaaacg ttctgtactg 900
tccaagtcag ttcctgggta caagccaaag gtcattccaa atgctatatg tggaatttgt 960
ctgaagggtg aggagtccaa caagaaagga aaggctgaat cacttatata ctgctcccaa 1020
tgtgagaata gtggccatcc ttcttgcttg gatatgacaa tggagcttgt ttctatgatt 1080
aagacctacc catggcagtg tatggaatgt aaaacatgca ttatatgtgg acaacccac 1140
catgaagaag aaatgatgtt ctgtgatatg tgtgacagag gttatcatac tttttgtgtg 1200
ggccttggtg ctattccatc aggtcgctgg atttgtgact gttgtcagcg ggcccccca 1260
acacccagga aagtgggcag aagggggaaa aacagcaaag agggataaaa tagtttttga 1320
ctctaatact gtatatgcat ttaagtggaa tatttggtgc catttacaac attattttca 1380
tgccaataaa agattttttt tgcaaattat gagcttaaaa tctgcagtta tttctgtaa 1440
aagtacgctt actctcgaaa ctaactccag gtagagaatt catcttcaa agtattttat 1500
agtaaccttg gctcactcca aaaattcagt ggaaatgttt agtaacttaa gataactaac 1560

```

```

tgtttctcca tagcccaaaa agttaatddd catgaaactt cctaactctac attgtttccg 1620
gcctaccata ggtagcactg acaaagttat ttaataactg aagaattttc atagggtatga 1680
caaatggccc actaagattt ggtgcagctg gatttagagt tgtcattatt ggactgggtac 1740
aggaacaaaac tttgtaaaata cctgcctgcc aggaaatcct ttttgatatag aaaagtacca 1800
tcacctactt ggggtacagg catgaggctt tagtccaggc tcagggaagt gtacgtaaaat 1860
catttccaac ttgatttttag taactcttga aaacttacac caacttcggg tagaatctcc 1920
agagtaaaat tacaaagtta tcaacctttt gattttgtgtc acagcatgaa aggtttgctct 1980
attttatata aacctgttac tgcaatcatt tttagtcaac ctgcctaata aaaatggagt 2040
ctaaacactt tgtgcacagt ctttttatag gaattatgat ctttaaaata ctgtgcttgc 2100
tgcttttccct atttttgggg taactgaggt aacaaaatgc gtatggcttt 2150

```

&lt;210&gt; 7

&lt;211&gt; 211

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 2059022

&lt;400&gt; 7

```

Met Ala Asn Ser Gly Leu Gln Leu Leu Gly Phe Ser Met Ala Leu
  1             5             10             15
Leu Gly Trp Val Gly Leu Val Ala Cys Thr Ala Ile Pro Gln Trp
             20             25             30
Gln Met Ser Ser Tyr Ala Gly Asp Asn Ile Ile Thr Ala Gln Ala
             35             40             45
Met Tyr Lys Gly Leu Trp Met Asp Cys Val Thr Gln Ser Thr Gly
             50             55             60
Met Met Ser Cys Lys Met Tyr Asp Ser Val Leu Ala Leu Ser Ala
             65             70             75
Ala Leu Gln Ala Thr Arg Ala Leu Met Val Val Ser Leu Val Leu
             80             85             90
Gly Phe Leu Ala Met Phe Val Ala Thr Met Gly Met Lys Cys Thr
             95            100            105
Arg Cys Gly Gly Asp Asp Lys Val Lys Lys Ala Arg Ile Ala Met
            110            115            120
Gly Gly Gly Ile Ile Phe Ile Val Ala Gly Leu Ala Ala Leu Val
            125            130            135
Ala Cys Ser Trp Tyr Gly His Gln Ile Val Thr Asp Phe Tyr Asn
            140            145            150
Pro Leu Ile Pro Thr Asn Ile Lys Tyr Glu Phe Gly Pro Ala Ile
            155            160            165
Phe Ile Gly Trp Ala Gly Ser Ala Leu Val Ile Leu Gly Gly Ala
            170            175            180
Leu Leu Ser Cys Ser Cys Pro Gly Asn Glu Ser Lys Ala Gly Tyr
            185            190            195
Arg Ala Pro Arg Ser Tyr Pro Lys Ser Asn Ser Ser Lys Glu Tyr
            200            205            210
Val

```

&lt;210&gt; 8

&lt;211&gt; 1546

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte clone 2059022

&lt;400&gt; 8

```

cccacgcgtc cgctcacctc cgagccacct ctgctgcgca ccgcagcctc ggacctacag 60
cccaggatag tttgggactt gccggcgctc agaaacgcgc ccagacggcc cctccacctt 120
ttgtttgcct agggtcgcgc agagcgcccg gaggaaccg cctggccttc ggggaccacc 180
aattttgtct ggaaccaccc tcccggcgta tccactccc tgtgcgcga ggccatcgct 240
tactggagg ggtcgatttg tgtgtagttt ggtgacaaga ttgcatcca cctggcccaa 300
accctttttg tctctttggg tgaccggaaa actccacctc aagttttctt ttgtgggct 360
gcccccaag tgcgtttgt tttactgtag ggtctccccg cccggcgccc ccagtgtttt 420
ctgaggcgcg aaatggccaa ttcgggcctg cagttgctgg gcttctccat ggccctgctg 480
ggctgggtgg gtctgggtgg ctgcaccgcc atcccgagc ggcatgatg ctctatgctg 540
ggtgacaaca tcatcacggc ccaggccatg tacaaggggc tgtggatgga ctgcgtcacg 600
cagagcacgg ggatgatgag ctgcaaaatg tacgactcgg tgctcgccct gtccgcggcc 660
ttgcaggcca ctgcagccct aatggtggtc tccctggtgc tgggcttcct ggccatgttt 720
gtggccacga tgggcatgaa gtgcacgcgc tgtgggggag acgacaaagt gaagaaggcc 780
cgtatagcca tgggtggagg cataattttc atcgtggcag gtcttgccgc cttggttagct 840
tgctctggt atggccatca gattgtcaca gacttttata accctttgat ccctaccaac 900
attaagtatg agtttgccc tgccatcttt attggtctgg cagggctctg cctagtcac 960
ctgggagggt cactgctctc ctgttctgt cctgggaatg agagcaaggc tgggtaccgt 1020
gcaccccgct cttaccctaa gtccaactct tccaaggagt atgtgtgacc tgggatctcc 1080
ttgccccagc ctgacaggct atgggagtgt ctgatgcct gaaagggcct ggggctgagc 1140
tcagcctgtg ggcagggtgc cggacaaagg cctcctggtc actctgtccc tgactccat 1200
gtatagtcct cttgggttgg ggggtggggg gtgcggttg tgggagagac aaaaagagg 1260
agagtgtgct tttgtacag taataaaaaa taagtattgg gaagcaggct ttttccctt 1320
cagggcctct gctttctcc cgtccagatc ctgacaggga gcttggaaac ttagtgacc 1380
tacttcagtt cagaacactt agcaccacac tgactccact gacaattgac taaaagatgc 1440
aggtgctcgt atctcgacat tcattccac cccctctta tttaaatagc taccaaagta 1500
cttctttttt aataaaaaaa taaagatttt tattaataaaa aaaaaa 1546

```

&lt;210&gt; 9

&lt;211&gt; 239

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;300&gt;

&lt;308&gt; P56198

&lt;400&gt; 9

```

Met Asp Tyr Ala Met Lys Ser Leu Ser Leu Leu Tyr Pro Arg Ser
  1             5             10             15
Leu Ser Arg His Val Ala Val Ser Thr Ala Val Val Thr Gln Gln
             20             25             30
Leu Val Ser Lys Pro Ser Arg Glu Thr Pro Arg Ala Arg Pro Cys
             35             40             45
Arg Val Ser Thr Ala Asp Arg Lys Val Arg Lys Gly Ile Met Ala
             50             55             60
His Ser Leu Glu Asp Leu Leu Asn Lys Val Gln Asp Ile Leu Lys

```

	65		70		75
Leu Lys Asp Lys	Pro Phe Ser Leu Val	Leu Glu Glu Asp Gly Thr			
	80		85		90
Ile Val Glu Thr	Glu Glu Tyr Phe Gln Ala	Leu Ala Lys Asp Thr			
	95		100		105
Met Phe Met Val	Leu Leu Lys Gly Gln Lys	Trp Lys Pro Pro Ser			
	110		115		120
Glu Gln Arg Lys	Lys Arg Ala Gln Leu Ala	Leu Ser Gln Lys Pro			
	125		130		135
Thr Lys Lys Ile	Asp Val Ala Arg Val Thr	Phe Asp Leu Tyr Lys			
	140		145		150
Leu Asn Pro Gln	Asp Phe Ile Gly Cys Leu	Asn Val Lys Ala Thr			
	155		160		165
Leu Tyr Asp Thr	Tyr Ser Leu Ser Tyr Asp	Leu His Cys Tyr Lys			
	170		175		180
Ala Lys Arg Ile	Val Lys Glu Ile Val Arg	Trp Thr Leu Phe Ser			
	185		190		195
Met Gln Ala Thr	Gly His Met Leu Leu Gly	Thr Ser Ser Tyr Met			
	200		205		210
Gln Gln Phe Leu	Asp Ala Thr Glu Glu Glu	Gln Pro Ala Lys Ala			
	215		220		225
Lys Pro Ser Ser	Leu Leu Pro Ala Cys Leu	Lys Met Leu Gln			
	230		235		

<210> 10  
 <211> 331  
 <212> PRT  
 <213> Homo sapiens

<300>  
 <308> g2065561

<400> 10  
 Met Glu Val Thr Gly Asp Ala Gly Val Pro Glu Ser Gly Glu Ile  
 1 5 10 15  
 Arg Thr Leu Lys Pro Cys Leu Leu Arg Arg Asn Tyr Ser Arg Glu  
 20 25 30  
 Gln His Gly Val Ala Ala Ser Cys Leu Glu Asp Leu Arg Ser Lys  
 35 40 45  
 Ala Cys Asp Ile Leu Ala Ile Asp Lys Ser Leu Thr Pro Val Thr  
 50 55 60  
 Leu Val Leu Ala Glu Asp Gly Thr Ile Val Asp Asp Asp Tyr  
 65 70 75  
 Phe Leu Cys Leu Pro Ser Asn Thr Lys Phe Val Ala Leu Ala Ser  
 80 85 90  
 Asn Glu Lys Trp Ala Tyr Asn Asn Ser Asp Gly Gly Thr Ala Trp  
 95 100 105  
 Ile Ser Gln Glu Ser Phe Asp Val Asp Glu Thr Asp Ser Gly Ala  
 110 115 120  
 Gly Leu Lys Trp Lys Asn Val Ala Arg Gln Leu Lys Glu Asp Leu  
 125 130 135  
 Ser Ser Ile Ile Leu Leu Ser Glu Glu Asp Leu Gln Met Leu Val

	140	145	150
Asp Ala Pro Cys	Ser Asp Leu Ala Gln	Glu Leu Arg Gln Ser	Cys
	155	160	165
Ala Thr Val Gln	Arg Leu Gln His Thr	Leu Gln Gln Val Leu	Asp
	170	175	180
Gln Arg Glu Glu	Val Arg Gln Ser Lys	Gln Leu Leu Gln Leu	Tyr
	185	190	195
Leu Gln Ala Leu	Glu Lys Glu Gly Ser	Leu Leu Ser Lys Gln	Glu
	200	205	210
Glu Ser Lys Ala	Ala Phe Gly Glu Glu	Val Asp Ala Val Asp	Thr
	215	220	225
Gly Ile Ser Arg	Glu Thr Ser Ser Asp	Val Ala Leu Ala Ser	His
	230	235	240
Ile Leu Thr Ala	Leu Arg Glu Lys Gln	Ala Pro Glu Leu Ser	Leu
	245	250	255
Ser Ser Gln Asp	Leu Glu Leu Val Thr	Lys Glu Asp Pro Lys	Ala
	260	265	270
Leu Ala Val Ala	Leu Asn Trp Asp Ile	Lys Lys Thr Glu Thr	Val
	275	280	285
Gln Glu Ala Cys	Glu Arg Glu Leu Ala	Leu Arg Leu Gln Gln	Thr
	290	295	300
Gln Ser Leu His	Ser Leu Arg Ser Ile	Ser Ala Ser Lys Ala	Ser
	305	310	315
Pro Pro Gly Asp	Leu Gln Asn Pro Lys	Arg Ala Arg Gln Asp	Pro
	320	325	330
Thr			

&lt;210&gt; 11

&lt;211&gt; 371

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;300&gt;

&lt;308&gt; g606661

&lt;400&gt; 11

Met Glu Gln Cys	His Asn Tyr Asn Ala Arg	Leu Cys Ala Glu Arg
1	5	10
Ser Val Arg Leu	Pro Phe Leu Asp Ser Gln	Thr Gly Val Ala Gln
	20	25
Ser Asn Cys Tyr	Ile Trp Met Glu Lys Arg	His Arg Gly Pro Gly
	35	40
Leu Ala Ser Gly	Gln Leu Tyr Ser Tyr Pro	Ala Arg Arg Trp Arg
	50	55
Lys Lys Arg Arg	Ala His Pro Pro Glu Asp	Pro Arg Leu Ser Phe
	65	70
Pro Ser Ile Lys	Pro Asp Thr Asp Gln Thr	Leu Lys Lys Glu Gly
	80	85
Leu Ile Ser Gln	Asp Gly Ser Ser Leu Glu	Ala Leu Leu Arg Thr
	95	100
Asp Pro Leu Glu	Lys Arg Gly Ala Pro Asp	Pro Arg Val Asp Asp
	110	115
Asp Ser Leu Gly	Glu Phe Pro Val Ser Asn	Ser Arg Ala Arg Lys
	125	130
		135

Arg Ile Ile Glu Pro Asp Asp Phe Leu Asp Asp Leu Asp Asp Glu		
	140	145
Asp Tyr Glu Glu Asp Arg Pro Lys Arg Arg Gly Lys Gly Lys Ser		150
	155	160
Lys Ser Lys Gly Val Ser Ser Ala Arg Lys Lys Leu Asp Ala Ser		165
	170	175
Ile Leu Glu Asp Arg Asp Lys Pro Tyr Ala Cys Asp Ile Cys Gly		180
	185	190
Lys Arg Tyr Lys Asn Arg Pro Gly Leu Ser Tyr His Tyr Ala His		195
	200	205
Ser His Leu Ala Glu Glu Glu Gly Glu Asp Lys Glu Asp Ser Arg		210
	215	220
Pro Pro Thr Pro Val Ser Gln Arg Ser Glu Glu Gln Lys Ser Lys		225
	230	235
Lys Gly Pro Asp Gly Leu Ala Leu Pro Asn Asn Tyr Cys Asp Phe		240
	245	250
Cys Leu Gly Asp Ser Lys Ile Asn Lys Lys Thr Gly Gln Pro Glu		255
	260	265
Glu Leu Val Ser Cys Ser Asp Cys Gly Arg Ser Gly His Pro Ser		270
	275	280
Cys Leu Gln Phe Thr Pro Val Met Met Ala Ala Val Lys Thr Tyr		285
	290	295
Arg Trp Gln Cys Ile Glu Cys Lys Cys Cys Asn Leu Cys Gly Thr		300
	305	310
Ser Glu Asn Asp Asp Gln Leu Leu Phe Cys Asp Asp Cys Asp Arg		315
	320	325
Gly Tyr His Met Tyr Cys Leu Thr Pro Ser Met Ser Glu Pro Pro		330
	335	340
Glu Gly Ser Trp Ser Cys His Leu Cys Leu Asp Leu Leu Lys Glu		345
	350	355
		360
Lys Ala Ser Ile Tyr Gln Asn Gln Asn Ser Ser		
	365	370

<210> 12  
 <211> 220  
 <212> PRT  
 <213> Homo sapiens

<300>  
 <308> g2570129

<400> 12

Met Ser Met Gly Leu Glu Ile Thr Gly Thr Ala Leu Ala Val Leu		
1	5	10
Gly Trp Leu Gly Thr Ile Val Cys Cys Ala Leu Pro Met Trp Arg		15
	20	25
Val Ser Ala Phe Ile Gly Ser Asn Ile Ile Thr Ser Gln Asn Ile		30
	35	40
Trp Glu Gly Leu Trp Met Asn Cys Val Val Gln Ser Thr Gly Gln		45
	50	55
Met Gln Cys Lys Val Tyr Asp Ser Leu Leu Ala Leu Pro Gln Asp		60
	65	70
Leu Gln Ala Ala Arg Ala Leu Ile Val Val Ala Ile Leu Leu Ala		75
	80	85
		90



Ala Phe Gly Leu Leu Val Ala Leu Val Gly Ala Gln Cys Thr Asn		
	95	100
Cys Val Gln Asp Asp Thr Ala Lys Ala Lys Ile Thr Ile Val Ala		105
	110	115
Gly Val Leu Phe Leu Leu Ala Ala Leu Leu Thr Leu Val Pro Val		120
	125	130
Ser Trp Ser Ala Asn Thr Ile Ile Arg Asp Phe Tyr Asn Pro Val		135
	140	145
Val Pro Glu Ala Gln Lys Arg Glu Met Gly Ala Gly Leu Tyr Val		150
	155	160
Gly Trp Ala Ala Ala Ala Leu Gln Leu Leu Gly Gly Ala Leu Leu		165
	170	175
Cys Cys Ser Cys Pro Pro Arg Glu Lys Lys Tyr Thr Ala Thr Lys		180
	185	190
Val Val Tyr Ser Ala Pro Arg Ser Thr Gly Pro Gly Ala Ser Leu		195
	200	205
Gly Thr Gly Tyr Asp Arg Lys Asp Tyr Val		210
	215	220

The first part of the paper discusses the importance of maintaining accurate records of all transactions. It is essential for the business to have a clear and concise record of all income and expenses, as this will be necessary for the preparation of the tax return. The second part of the paper discusses the importance of maintaining accurate records of all assets and liabilities. This is also essential for the preparation of the tax return, as it will be necessary to report the value of all assets and liabilities. The third part of the paper discusses the importance of maintaining accurate records of all tax payments. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax payments. The fourth part of the paper discusses the importance of maintaining accurate records of all tax credits. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax credits. The fifth part of the paper discusses the importance of maintaining accurate records of all tax deductions. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax deductions. The sixth part of the paper discusses the importance of maintaining accurate records of all tax credits and deductions. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax credits and deductions. The seventh part of the paper discusses the importance of maintaining accurate records of all tax credits and deductions. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax credits and deductions. The eighth part of the paper discusses the importance of maintaining accurate records of all tax credits and deductions. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax credits and deductions. The ninth part of the paper discusses the importance of maintaining accurate records of all tax credits and deductions. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax credits and deductions. The tenth part of the paper discusses the importance of maintaining accurate records of all tax credits and deductions. This is also essential for the preparation of the tax return, as it will be necessary to report the amount of all tax credits and deductions.